



**CGIAR Research Program on  
Climate Change, Agriculture and Food Security (CCAFS)**

**Introduction to Dataverse**  
*Video Transcript*

---

**October 2013**



Data Management Guidelines by [Statistical Services Centre, University of Reading](http://www.statistics.ac.uk/) is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License](https://creativecommons.org/licenses/by-nc-sa/3.0/).

Permissions beyond the scope of this license may be available at [www.reading.ac.uk/ssc](http://www.reading.ac.uk/ssc).

These materials were produced for and with funding from the Climate Change Agriculture and Food Security Research Program of the Consultative Group on International Agricultural Research (CGIAR).

## Introduction

The Dataverse Network is an open source application for publishing, sharing, referencing, extracting and analysing research data. It is created by the Institute for Quantitative Social Science at Harvard University. Although aimed originally at social science data, a Dataverse can be created to hold research data on any subject. For example the World Agroforestry Centre (ICRAF) has a Dataverse currently with 22 studies.

An introductory video is available on the home page of the Dataverse network project at [thedata.org](http://thedata.org) – if you haven't already done so we recommend you view this short video and download the fact sheet before continuing here.

## Citation

One thing that often stops researchers sharing their data is the belief that they won't get the credit for the work they have done. With Dataverse a unique citation is automatically generated when a study is created and this should be used as a reference to the study. The citation has 6 components: the author(s), title, year and distributor are the human readable elements, of these only the title is compulsory though author and year are recommended. The 2 machine readable elements are: a unique global identifier (often called a "handle") beginning with "hdl", and a universal numeric fingerprint (UNF).

Here we see an example of a citation for a study on the SSC Dataverse. In this instance we don't have author, year or distributor so this just includes the title as the human-readable element.

For more information about citations go to the Data Citation page of the Dataverse Network website.

## Hosting

There are two options for hosting your Dataverse. The easiest is to have your Dataverse hosted at Harvard. This is a free service to all researchers. The infrastructure at Harvard is very good and great support is available to help you set up your Dataverse – you can also include your own branding in the header including working links so that your Dataverse has the look and feel of your own website. For example the Statistical Services Centre Dataverse has live links in the header section taking you to the SSC website.

The other option is self-hosting and you can download and install the relevant software. This gives you more administrative control but you would need an IT expert to install and manage the site including upgrading, taking backups etc. You would also need good server infrastructure for hosting the application.

We would generally recommend the first option – i.e. having the Dataverse hosted at Harvard. To find out more about the two options follow the links to "As a Service" and "As a software package" on the Dataverse Network website.

## Structure of a Dataverse

Each Dataverse contains studies and collections of studies and each study contains cataloguing information describing the data plus the data files themselves and any associated files (e.g. reports, results files, study guidelines, etc.)

Collections are optional – you can have all your studies at the same level in your Dataverse – this is fine if you have just a few studies but as more and more studies are added you may find it useful to divide them into collections. Think of this in terms of how you might divide your files into folders on your PC.

For example in the SSC Dataverse there are 5 studies – one is the Targeted input programme (TIP) and the others are for the Starter Pack Evaluation Programme (Starter Pack 2). Within the Starter Pack there are 3 modules so we have the top level which we have made a collection which contains the 3 modules. Our archive files on our local drive reflect this structure with a top level folder for Starter Pack 2 and 3 sub-folders for the modules.

On their Dataverse ICRAF have created collections for each region.

## Permissions

When a study is released the default is for public access, but you can choose to restrict the entire study giving access to named users only – or you can restrict individual files within a study even if the study itself has public access. For example we can see here on the ICRAF Dataverse there are some studies that are restricted as indicated by the padlock symbol; on the CCAFS Dataverse the Baseline Household Survey study is public access but the data files are restricted.

## Summary

Dataverse is primarily an archiving facility. However, you can create a Dataverse and start creating studies early on in your project, gradually building the archive as each stage of the project is completed. Studies are only made public once they are released, so you can continue building your archive until you are ready to release it. If you have a Data and document store for your project files, then creating your archive should be relatively straight-forward as you can keep the same structure.

This video is just a brief introduction to the concept – later videos will show how to start creating your Dataverse.

## Appendix I – CCAFS Data Management Support Pack

This document is part of the CCAFS Data Management Support Pack produced by the Statistical Services Centre, University of Reading, UK. The following materials are available in the pack:

0. Data Management Strategy
  - a. CCAFS Data Management Strategy
1. Research Protocols
  - a. Writing Research Protocols – a statistical perspective
  - b. Preparation of Research Protocols – Good Practice Case Study
  - c. What is a Research Protocol, and how to use one (Video & Transcript)
  - d. Details of what a Research Protocol should contain (Video & Transcript)
2. Data Management Policies & Plans
  - a. Creating a Data Management Plan
  - b. Data Management Plan (Video & Transcript)
  - c. Example Data Management Activity Plan
  - d. Example Consent Form
3. Budgeting & Planning
  - a. Budgeting & Planning for Data Management
  - b. ToR Data Support Staff
  - c. Budgeting & Planning (Video & Transcript)
4. Data Ownership
  - a. Data Ownership and Authorship
  - b. Template – Data Ownership Agreement
  - c. CCAFS Data Ownership & Sharing Agreement
  - d. Data Ownership & Authorship (Video & Transcript)
5. Data & Document Storage
  - a. Creating and Using a DDS
  - b. DDS Introduction – (Video & Transcript)
  - c. DDS Organisation – (Video & Transcript)
  - d. DDS Ownership – (Video & Transcript)
  - e. Introduction to Dropbox – (Video & Transcript)
6. Archiving & Sharing
  - a. Archiving & Sharing Data
  - b. Data and Documents to Submit for Archiving – a checklist
  - c. MetaData
  - d. Archiving & Sharing (Video & Transcript)
  - e. Metadata (Video & Transcript)
  - f. CCAFS HBS Questionnaire
  - g. CCAFS HHS Code Book
  - h. CCAFS Training Manual for Field Supervisors



7. CCAFS Data Portals
  - a. Portals for CCAFS Outputs
  - b. AgTrials Summary
  - c. CCAFS-Climate Summary
  - d. DSpace Introduction
  - e. Introduction to Dataverse (Video & Transcript)
  - f. Creating a Dataverse (Video & Transcript)
  - g. Dataverse Study Catalogue
  - h. CCAFS Dataverse (Video & Transcript)
8. Data Quality & Organisation
  - a. Data Quality Assurance
  - b. Guidance for handling different types of Data
  - c. Transition from Raw to Primary Data
  - d. Data Quality Assurance (Video & Transcript)
  - e. Guidance for handling different types of data (Video & Transcript)
  - f. Transition from Raw to Primary Data (Video & Transcript)